Ellis Godard
Statistics Handout
Subgroup Comparisons

Example of Using Subgroups to Make Bivariate Comparisons

This data, from the 1992 National Longitudinal Survey of Youth, shows that the subgroup of white men and subgroup of black men have basically the same mean, median, and mode for *age* — the same normal curve, with the same central tendency.

A. White Men

Variable	Mean	Median	Mode
Age (Years)	30.88	31	29
Hourly Wage ($)	12.36	11.51	–
Family Yearly Income ($)	38,345	36,125	–

B. Black Men

Variable	Mean	Median	Mode
Age (Years)	30.87	31	29
Hourly Wage ($)	9.75	8.50	–
Family Yearly Income ($)	27,799	23,000	–

However, the distributions of *income* are neither normal nor the same for the two subgroups. The black men report a mean hourly wage of only 78.9% of the mean hourly wage reported by white men (9.75 vs. 12.36).

Moreover, while both distributions are skewed to the right (the mean is higher than the median, indicating a lower concentration but some outliers on the high end), the difference is much more pronounced for blacks (with a mean/median ratio of 1.15 for black men, as opposed to 1.07 for white men) which suggests that their skew is even larger — their outliers are even further from the rest of the group, and their wages are even more suppressed in the lower tier than the measures of central tendency suggest.

Limits to this Comparison

The inferences I've made about dispersion are not wholly warranted. And my implication that the differences observed are big enough to matter, are really just ad hoc assertions. They hold no statistical weight. What we really need here are statistical tests, including a t-test to assess whether the Difference of Means is statistically significant, as well as perhaps an ANOVA (or some variant on an F-test) in order to assess whether the variances are as different as I've argued.